Personal Autonomy

First published Tue May 28, 2002; substantive revision Thu Feb 15, 2018

Autonomous agents are self-governing agents. But what is a self-governing agent? Governing oneself is no guarantee that one will have a greater range of options in the future, or the sort of opportunities one most wants to have. Since, moreover, a person can govern herself without being able to appreciate the difference between right and wrong, it seems that an autonomous agent can do something wrong without being to blame for her action. What, then, are the necessary and sufficient features of this self-relation? Philosophers have offered a wide range of competing answers to this question.

1. Introduction

When people living in some region of the world declare that their group has the right to live autonomously, they are saying that they ought to be allowed to govern themselves. In making this claim, they are, in essence, rejecting the political and legal authority of those not in their group. They are insisting that whatever power these outsiders may have over them, this power is illegitimate; they, and they alone, have the authority to determine and enforce the rules and policies that govern their lives.

When an individual makes a similar declaration about some sphere of her own life, she, too, is denying that anyone else has the authority to control her activity within this sphere; she is saying that any exercise of power over this activity is illegitimate unless she authorizes it herself. Most of the reasons that can be offered in support of this claim have correlates in the case of demands for group autonomy. But there is one very important exception: a reason that takes us beyond politics, to the metaphysics of agency.

An agent is one who acts. In order to act, one must initiate one’s action. And one cannot initiate one’s action without exercising one’s power to do so. Since nothing and no one has the power to act except the agent herself, she alone is entitled to exercise this power, if she is entitled to act. This means that insofar as someone is an agent, i.e., insofar as she is one who acts—she is correct to regard her own commitments to acting, her own judgments and decisions about how she should act, as authoritative. Indeed, if she were to challenge the authority that is an essential feature of her judgments and decisions, then they would cease to be her own practical conclusions. Their power to move her would cease to be a manifestation of her power to move herself; it would not be the power of her own agency.

In short, every agent has an authority over herself that is grounded, not in her political or social role, nor in any law or custom, but in the simple fact that she alone can initiate her actions. To be sure, it might be unwise for someone to follow the commands she gives to herself when she “makes up her mind.” The point, however, is that she has no conceivable option. In order to form an intention to do one thing rather than another, an agent must regard her own judgment about how to act as authoritative—even if it is only the judgment that she should follow the command or advice of someone else. This tight connection between being an agent and having authority has no correlate in cases where the authority at issue is political. Anyone can coherently (and often plausibly) challenge the political authority of
Despite the special inalienable nature of our authority over ourselves, it is possible for us to fail to govern ourselves, just as it is possible for a political leader to fail to govern those who fall within her domain. Indeed, precisely because our authority over our own actions is an intrinsic feature of our agency, our deference to this authority is but the form of self-government. It is no guarantee that whenever we act, the forces that move us owe their power to our power to decide what to do. Just as a political leader’s official status is compatible with her having no real power to call the shots, so too, a person can have an authoritative status with respect to her motives without having any real power over them. Though it is an agent’s job to determine how she will act, she can do this job without really being in control. Of course, no one can govern herself without being subject to influences whose power does not derive from her own authority: everything we do is a response to past and present circumstances over which we have no control. But some of the forces that move us to act do not merely affect which actions we choose to perform, nor how we govern ourselves in making these choices. They influence us in a way that makes a mockery of our authority to determine our own actions. They undermine our autonomy.

What distinguishes autonomy-undermining influences on a person’s decision, intention, or will from those motivating forces that merely play a role in the self-governing process? This is the question that all accounts of autonomy try to answer. As the number and variety of these accounts indicate, the distinction is extremely elusive. There is certainly widespread agreement about the paradigm threats to personal autonomy: brainwashing and addiction are the favorite examples in the philosophical literature. But philosophers seem unable to reach a consensus about the precise nature of these threats. They cannot agree about how it is that certain influences on our behavior prevent us from governing ourselves.

This disagreement about the defining characteristics of autonomous agency reflects the fact that even as concrete examples appear to call attention to a very real difference between those who govern themselves and those who do not, there are significant conceptual obstacles to making sense of this distinction. These obstacles are tied to the very feature of agency mentioned above—the feature that appears to support the demand that individuals be granted considerable political and legal power. If an agent fails to govern herself when she acts, this must be because what she does is independent of her power to determine how she will act. But if she necessarily has the authority to determine how she will act, and if this essential feature of agency is inseparable from the fact that she necessarily defers to herself whenever she initiates her action, then how can her behavior possibly escape her control? Intuitively, an agent can fall under the sway of desires, or urges, or compulsions whose power is at odds with her own power as an agent; she can be moved by such impulses “in spite of herself.” But in what sense, exactly, are such motives “external” to the agent herself? How can their power to move her fail to be a manifestation of her power to act? How can their power reduce her authorization of her action to a mere formality? It is difficult to answer these questions when the governing agent and the agent she governs are one and the same.

(Again, the perplexity to which these questions give voice does not have a correlate in the political case. We can easily grasp the idea of a country’s army (or legislative body, or cabinet ministers) dictating to the president what legislation he must approve; for in this case there are (at least) two independently identifiable decision-makers—each with its own point of view, each with its own power. The difficulty in the case where the relevant powers are all within the psyche of a single individual agent is that there is no such independently identifiable pair of standpoints in terms of which we can distinguish the powers that bully this agent from the powers that can be attributed to the agent herself. An account of the conditions under which an individual agent is bullied by her motives is, at the same time, an account of what makes a motive external to the agent’s own standpoint.)

2. Four More or Less Overlapping Accounts of Personal Autonomy

Philosophers have proposed many different accounts of the autonomous agent’s special relation to her own motives. According to one prominent conception, which one might call “coherentist,” an agent governs her own action if and only if she is motivated to act as she does because this motivation coheres with (is in harmony with) some mental state that represents her point of view on the action. The relevant mental state varies from account to account. According to one popular story, an agent’s point of view is constituted by her highest-order desires regarding which of her first-order desires moves her to act. (Frankfurt 1988c) According to another story, her point of view is constituted by her (contemporaneous or long-term) evaluative judgments regarding which actions are (most) worth performing. (Watson 1975) Still another account adds that there must also be harmony between what the agent does and her more or less long-term plans (Bratman 1979 and 2007). And others appeal to the relatively stable network of emotional states constitutive of “caring” (Frankfurt 1989f and 1999d, Jaworska, Shoemaker 2003) or to the agent’s character traits (Dworkin, R.), or to her most thoroughly “integrated” psychological states (Arpaly and Schroeder).

All these accounts reflect the intuition that an action cannot be attributed to the agent herself if, even as she performs this action, she occupies a point of view from which she repudiates what she is doing. More carefully, such an action cannot be the agent’s doing in the way that it must be if it is to qualify as an instance of self-government. According to this intuition, if someone repudiates, or in some other way dissociates herself from, the causal efficacy of her own motives, then the power of these motives is independent of her authority. If, on the other hand, she endorses these motives, whether implicitly or explicitly, then her actions occur with her permission, if not necessarily at her command. Under what conditions does someone count as endorsing or repudiating her motives? Each account offers a different
Not only does there appear to be a tight conceptual connection between self-governing agency and synchronic psychic unity; there also appears to be a connection between self-governing agency and the diachronic unity of one’s later self with one’s earlier self. This is the connection central to accounts that identify self-governing agents with agents constrained by plans, or by well-integrated emotions, or traits of character. Agents persist through time; and so, these accounts stress, an agent’s point of view is not simply a function of whatever mental state(s) she happens to be in at some point in time. Because an agent’s plans play a crucial role in ensuring that she is more than a mere collection or sequence of mental states, it is reasonable to think that whether her motives have her support depends on whether they are constrained by these plans. So, too, it is reasonable to think that her stance toward her motives is determined by her long-term values and/or her relatively stable commitments and cares.

On a strict coherentist conception of autonomy, autonomous agents can be moved by desires they are helpless to resist: though an addict fails to govern herself if she would rather resist her irresistible urge to take drugs, she is an autonomous agent if she has no objection to her addiction and its motivational effects. According to the coherentist, moreover, both the origin and the content of a person’s higher-order attitudes (evaluative judgments, plans) are irrelevant to whether she is an autonomous agent. She need have done nothing to bring it about that she has these attitudes; and the attitudes need not be especially rational or well-informed. Coherentist accounts are thus doubly internalist. They express the intuition that whether we govern ourselves depends on neither how we came to be to who we are (a fact that is prior to (and in this sense external to) the action itself) nor how our beliefs and attitudes relate to reality (a fact that is independent of (and in this sense external to) the beliefs and attitudes themselves). In other words, on these accounts, there need be no special relation between our autonomy-constituting attitudes and either the past circumstances that caused these attitudes or the present circumstances in response to which they move us to act.

Other accounts of autonomy introduce conditions that are externalist in one or both of these ways. According to those who advocate a reasons-responsive conception of autonomous agency, an agent does not really govern herself unless her motives, or the mental processes that produce them, are responsive to a sufficiently wide range of reasons for and against behaving as she does. (Fischer and Ravizza, Nelkin, Wolf) On accounts of this type, an agent who is unresponsive to the reasons for “standing behind,” or “backing up,” certain motives and not others is not in the proper position to authorize her own actions. Whether the relevant reasons are grounded in facts about her own desires and interests, or whether they have some independent source, the idea is that someone is not qualified to govern herself if she cannot understand what she (really) has reason to do, or (if this is a distinct handicap) is incapable of being moved by these reasons. In effect, her exercise of authority is so ill-conceived that it is powerless to confer legitimacy on her motives.

The feature of these accounts that most distinguishes them from coherentist accounts is the importance they attribute to an agent’s ability to appreciate the reasons she has. (Once she appreciates these reasons, her inability to act accordingly is, essentially, the inability to conform her act to her own judgment, and to her corresponding (higher-order) desire.) What, exactly, is the connection supposed to be between being out of touch with (evaluative and/or nonevaluative) reality and failing to govern oneself? Clearly, a person who fails to appreciate a wide range of reasons for action is unlikely to govern herself well: she is likely to do things that will, in the long run, thwart her own purposes and interests. The reasons-responsiveness conception of autonomy thus appears to reflect the intuition that when we do something very poorly, we do not really do it at all. There is, however, another possible underlying rationale for regarding ignorance as a threat to self-government. If doing Y is constitutive of doing Z, then if I authorize myself to be moved by the desire to do Y because I mistakenly believe that doing Y is a way of not doing Z, then there is an obvious sense in which I have not authorized myself to do what I am now doing when I am moved by the desire to do Y. So, if I have a general desire to do what is right and prudent, or, even more generally, a desire to do what I can justify to myself (and others), or, more generally still, a desire to be responsive to reasons, then insofar as I am moved to act in ways that are, in fact, incompatible with satisfying these desires, there is a sense in which I—who am committed to doing only what I have good (enough) reason to do—have not really authorized my action. Alternatively, we could say that, under these circumstances, something external to my power to guide myself by reasons has prevented me from exercising this power, and so has prevented me from governing myself.

An additional source of support for the reasons-responsive conception of autonomy comes from the thought that someone who cannot respond to the reasons there are must have a limited ability to reason. This brings us to a third popular approach to autonomous agency—an approach that stresses the importance of the reasoning process itself (Christman 1991, 1993 and Mele 1993, 1995). According to responsiveness-to-reasoning accounts, the essence of self-government is the capacity to evaluate one’s motives on the basis of whatever else one believes and desires, and to adjust these motives in response to one’s evaluations. It is the capacity to discern what “follows from” one’s beliefs and desires, and to act accordingly. One can exercise this capacity despite holding false beliefs of all kinds about what one has reason to do. Accordingly, on these accounts, being autonomous is not the same thing as being guided by correct evaluative and normative judgments.

The emphasis on an autonomous agent’s responsiveness to her own reasoning reflects the intuition that someone whose education consisted of a method of indoctrination that deprived her of the ability to call her own attitudes into question would, in effect, be governed by her “programmers,” not by herself. So, too, someone whose practical reasoning was directly manipulated by others would not govern herself by means of this reasoning. And so, it seems, she would have no power over the motives that this reasoning produced.
Like the coherentsists, advocates of responsiveness-to-reasoning accounts believe that the key to autonomous agency is the ability to distance oneself from one’s attitudes and beliefs—to occupy a standpoint that is not constituted by whatever mental states are moving one to act. They agree that motives authorized from this reflective standpoint are internal to the agent herself in a way that her other motives are not. Unlike the coherentsists, however, the reasoning-responsive theorists believe that there is more to the capacity for self-reflection than the capacity to hold higher-order attitudes. The authority of our higher-order attitudes is grounded, they claim, in the authority of the practical reasoning that supports these attitudes. So a self-governing agent does not merely endorse her motives: her endorsements are implicit claims about which motives have the support of her reason.

This fact is closely tied to another. Like many accounts that stress an autonomous agent’s responsiveness to reasons, responsiveness-to-reasoning accounts often suggest that self-government requires the capacity for self-transformation. On this assumption, an autonomous agent is someone who can change her mind when she discovers good reason to do so. In contrast, strict coherentsists insist that it is possible to act autonomously while being moved by desires that are not only irresistible when they produce their effects, but so integral to one’s identity that one could not possibly will to resist them. The conception of autonomous agency as responsiveness to reasoning clearly has a more internalist character than the conception of autonomous agency as responsiveness to reasons: according to those who stress the autonomous agent’s ability to evaluate her own motives, what counts is not the relation between the agent’s attitudes and external reality, but her ability to draw inferences from what she wants and believes, and by so doing, to reconsider—to rationally reflect upon—her other desires and beliefs. Insofar, however, as a responsiveness-to-reasoning account presupposes a particular conception of practical reasoning, it appeals to standards, or principles, that the agent herself might misapply, or fail to recognize altogether. Moreover, even if advocates of autonomy as responsiveness-to-reasoning have nothing in particular in mind when they speak of the process of “reflection,” “rational evaluation,” etc., reasoning is a norm-governed process that an agent might reject for reasons of her own. Responsiveness to reasoning accounts thus contain an externalist element that is absent from strict coherentsist accounts. They imply that an agent can be mistaken about whether she is really reasoning—and so can be mistaken about whether the power of her motives reflects the fact that she has the authority to determine her own actions.

This weak externalism naturally expands into more robust varieties. In particular, it supports the idea that whether an agent’s reasoning is really her way of governing her actions depends on which forces exert a nonrational influence on this reasoning. Even when indoctrination and other more or less imaginary forms of “mind control” do not prevent a person from reaching evaluative conclusions about her own motives, they can prevent her from thinking for herself. So, too, it seems, someone in the grip of compulsion or addiction can be so dominated by this condition that whatever facts she considers, and whatever conclusions she draws, cannot legitimately be attributed to her. One way to interpret these cases is to say that the person’s reasoning falls so far short of the norms of “rational reflection” that she is not really reasoning at all. Alternatively, one can say that her reasoning does not guarantee her autonomy because it is under the control of external forces.

Insofar as accounts of autonomy simply stipulate that certain influences on an agent’s intention-forming process “interfere with,” or “pervert,” this process, these accounts are incomplete. For they leave it mysterious why certain influences, and not others, are a threat to self-government. One response to this challenge is offered by reasons-responsive accounts: according to this response, the autonomy-undermining influences are the ones that prevent the reasoning process from being sufficiently sensitive to the reasons there are. Another – compatible – response appeals to the “relational” aspects of autonomy: other agents can prevent someone’s reasoning from qualifying as a mode of self-government by preventing the reasoner from developing the self-respect and/or self-trust necessary for forming a point of view that is truly her own (Benson 1994, 2000, Mackenzie and Stoljar, Anderson and Honneth). If an agent’s point of view does not reflect her respect for herself and her ability to set her own ends and assess the reasons relevant to pursuing some ends and not others, then the direction her reasoning takes cannot be attributed to her. (Though relational accounts of autonomy highlight the extent to which an agent’s capacity to govern herself depends on her interactions with other agents, it is important to note that self-respect and self-trust can also be undermined by experiences or psychological conditions that do not involve the actions of anyone else.)

A fourth conception of personal autonomy offers a very different response to the challenge of distinguishing (i) the determining causes that prevent an agent from governing herself when she employs her reason from (ii) the causes that determine how an agent governs herself when she reasons. According to this incompatibilist conception, each of these influences undermines the agent’s autonomy; cases of mind control simply call our attention to the fact that whenever our motives are causally determined by events over which we have no control, their power does not reflect our authority. (Pereboom) According to incompatibilists, if our actions can be fully explained as the effects of causal powers that are independent of us, then even if our beliefs and attitudes are among these effects, we do not govern them, and so we do not govern ourselves. (Kane 1996 and van Inwagen 1983)

Incompatibilist accounts of autonomy take many subtly different forms. So do the three other (compatibilist) accounts mentioned here. Some of the differences reflect disagreements over the extent to which the relevant conditions—coherence among higher- and lower-order attitudes, responsiveness to reasons, responsiveness to reasoning, freedom from determination by external causes—must actually obtain when an agent determines her will, or whether it is enough that under certain specified circumstances the agent would relate to her motives in the stipulated manner. There is also a difference of opinion about the scope of the relevant capacities: Must an autonomous agent be capable of responding to a wide range of reasons for and against her action? or is it enough that her motives are responsive to the “strongest,”...
may conclude that even though she has an overriding reason to give up her child for adoption, she cannot recognize contemporaneous normative verdicts without defying anything very deep about themselves. A woman, for example, compatible with autonomous agency.

The possibility of weakness of will points to the more general fact that an agent’s authorization of her own motives need not be just a matter of her approval of what she does. Rather, her accountability is grounded in her contemporaneous relation to what she is doing. Nonetheless, even under these conditions, the desires that move her to act do so on her own account. She acts for a reason that she herself deems adequate; and so she is not (adequately) governed by the desires that move her to act. It is notoriously difficult to make sense of such an exercise of autonomy, given that it is not clear how she exercises this authority, and thereby strengthens her motives for performing it.\[13\]

This way of interpreting the link between autonomous agency and responsiveness to reasons raises larger questions about the relationship between our practical impulses and our reason. The answers to these questions, and to those mentioned above, can be combined in many different ways. Not only, moreover, can each approach thus take a wide variety of forms, but the approaches themselves can (and often do) figure together as necessary or sufficient conditions in a single complex account.

3. Challenges to Identifying the Minimal Conditions of Personal Autonomy

All the proposals just considered contribute to our understanding of the various roles that agents can play in their own actions. They articulate various ideals that agents can realize to various degrees when they act. In so doing, they shed light on how, with the proper training, a very young child, whose deference to the authority of her own judgments is little more than the form of self-government, can develop into an exemplary self-governing agent.

This is a very important contribution. Nonetheless, it falls short of giving us everything we have reason to expect from an account of personal autonomy. In particular, challenges to the different approaches sketched above suggest that they do not spell out the minimal conditions under which a person’s exercise of authority over how she behaves reflects her own power to determine how she exercises this authority. Minimal self-government seems to require nothing more nor less than being the power behind whatever reasoning directly gives rise to one’s behavior. Yet none of the accounts we have canvassed here seems to capture this important, most basic, form of self-government. Nor, it seems, does any combination of these accounts.

The worry that the coherence of one’s contemporaneous attitudes does not suffice for even minimal self-government is grounded in the apparent possibility that a person could be brainwashed, or otherwise compelled, to endorse a given motive. Indeed, her brain could be manipulated in such a way that each of her endorsements is highly responsive to reasons. This has led some philosophers to supplement coherentist accounts of autonomy with additional conditions that place constraints on the causal history of an agent’s endorsements, constraints of the very sort singled out in the responsiveness-to-reasons accounts. These supplements face a significant challenge, however: it is very difficult to spell out the distinction between autonomy-conferring reasoning and autonomy-undermining reasoning without implicitly appealing to the very phenomenon one is trying to explain.

Even if historical approaches to autonomy can successfully overcome this difficulty, they do nothing to address the fact that coherence is not even necessary for autonomous agency. An agent need not sacrifice her autonomy in order to decide to act contrary to her long-term commitments and concerns; acting “out of character” is not a sufficient condition for failing to govern oneself. Though a “weak-willed” agent is hardly a paradigm example of someone who governs herself when she acts, she too plays a decisive role in the relative power of her own motives; she authorizes her behavior, even though she believes that she has good reason to act otherwise. It is notoriously difficult to make sense of such an exercise of authority.\[14\] For our purposes here, however, it suffices to note that if weakness of will is a genuine phenomenon, then human agents have the capacity to govern themselves in a way that they themselves take to be unjustified. They can assert an authority over themselves that challenges the authority of their very own reason.

Of course, someone whose action is caused in this way does not govern herself as thoroughly as someone whose will is “strong”; she acts for a reason that she herself deems inadequate; and so she is not (adequately) governed by the norms of her own thought. Nonetheless, even under these conditions, the desires that move her to act do so on her own authority. To use what is perhaps the most important metaphor in the literature on personal autonomy, the weak-willed agent “identifies with” her motives in whatever way she must in order to be accountable for their effects. It is not simply that what she does is the result of an earlier autonomous action. Rather, her accountability is grounded in her contemporaneous relation to what she is doing.

The possibility of weakness of will points to the more general fact that an agent’s authorization of her own motives need not take the form of the judgment that it would not be better to act otherwise. Various forms of perversity are perfectly compatible with autonomous agency. Some philosophers believe that it is also possible for agents to defy their own contemporaneous normative verdicts without defying anything very deep about themselves. A woman, for example, may conclude that even though she has an overriding reason to give up her child for adoption, she cannot recognize...
passivity, this diagnosis would be problematic. For it assimilates the addict to someone whose behavior does not even
are, in effect, “passive bystanders” to their own motives. But even if we could find a satisfactory account of the relevant
who fit this description are merely weak-willed, not all of them are: some unwilling addicts are not self-governing in even
though she would rather resist the motivating force of her addiction. It is widely agreed that, even if many people
Consider, for example, an alleged paradigm case of an agent who fails to govern herself: a person who takes drugs
autonomous
the conditions under which one is sufficiently self-governing to be accountable for the motivating power of one's desires.
account of what is special about self-governing
fails to govern her motives in the minimal way necessary to be accountable for them. It thus cannot be the key to any
whose power can be attributed to the agent herself from motives that are not in this sense internal. For if a desire
attributed to every potentially self-governing agent, it does not seem to be an adequate basis for distinguishing motives
a self-governing
that no agent can be alienated is the desire to have sufficient power to determine one’s own motives—the desire to be
evaluative judgment, or plan, or even an integrated combination of such attitudes. The only attitude from which it seems
reasons. Unfortunately, however, these accounts have problems of their own. Most importantly, it seems as though a
so, then this might seem to be a reason to favor accounts that tie autonomous agency to the agent’s responsiveness to
values that provide the underlying rationale for these plans.
Reflections along these lines have led some to conclude that we are bound to come up empty-handed as long as we
think of an agent's identification with her motives as a self-relation she is responsible for securing. For, as long as we
take this approach, we appear to be stuck with the question: under what conditions does the agent govern her
identification with some motive? what conditions must she satisfy in order to identify with the motives that move her to
identify with some of her motives and not others? under what conditions does she authorize the attitudes and/or mental
activities that issue in a given lower-order authorization? If we are to escape the regress such questions evoke, it seems
that there must be an attitude that can be identified with the agent’s point of view simply by virtue of being the attitude it is;
there must be an attitude from which no agent can possibly be alienated. To be sure, if—like most animals—racial agents
could not distance themselves from their own motives, then they would be incapable of governing themselves.
Self-government requires two points of view: that of the governing authority, and that of the governed. Nonetheless, if
there is to be an end to the potential regress of identifications, it seems that there must also be a limit to the capacity for
self-alienation.
As we have seen, no such limitation seems to apply where the mental state at stake is an agent's highest-order desire,
evaluative judgment, or plan, or even an integrated combination of such attitudes. The only attitude from which it seems
that no agent can be alienated is the desire to have sufficient power to determine one’s own motives—the desire to be a
self-governing agent. Even if, however, we leave to one side the question of whether this desire can really be attributed to
every potentially self-governing agent, it does not seem to be an adequate basis for distinguishing motives whose power can be attributed to the agent herself from motives that are not in this sense internal. For if a desire underlies every action performed by a potentially self-governing agent, then it plays a causal role even when an agent fails to govern her motives in the minimal way necessary to be accountable for them. It thus cannot be the key to any account of what is special about self-governing agency.
Perhaps there is no attitude to which we can point in order to distinguish between cases in which the power of an
agent’s motives can be directly attributed to her and cases in which her authority over her motives is a mere formality. If
so, then this might seem to be a reason to favor accounts that tie autonomous agency to the agent’s responsiveness to
reasons. Unfortunately, however, these accounts have problems of their own. Most importantly, it seems as though a
person can govern herself even if she does not understand the significance of what she is doing. To be sure, if
someone’s ignorance is perfectly reasonable, then she may not be to blame if she does something wrong. But under
such circumstances, what frees her from blame is the fact that she has good reason to be ignorant. There appears to
be no basis for assuming that, in addition to lacking certain relevant information, she is not really the (ill-informed)
power behind her authorization of her action.
In killing Desdemona, Othello fails to accomplish his aim of doing what he has good reason to do. But this does not
prevent him from being the author of his own actions. Nor would he necessarily have been prevented from governing
himself if, given his character and circumstances, he had been unable to “track” the evaluative and nonevaluative facts:
he would still have been accountable for what he did if his reason for doing it had been that he was too jealous, or too
stubborn, or too vain, or too hot-tempered to be capable of responding to the wide range of reasons against believing
that his wife was unfaithful to him—and the wide range of reasons against killing her even if she was unfaithful.
Carefully, to insist that he would have been accountable under these circumstances, we must abandon the
assumption that autonomous agency is possible even if all actions can, in principle, be explained in terms of
deterministic laws of nature. In other words, we must accept the incompatibilist thesis that if a person’s character is the
product of forces over which he never had any control, and if his character traits determine his choices, then even if his
motives are responsive to reasons, he is not responsible for their motivating force.
The preceding reflections call attention to how difficult it is to distinguish the conditions of ideal self-government from
the conditions under which one is sufficiently self-governing to be responsible for the motivating power of one’s desires.
Difficulty is manifested in the fact that as soon as we try to pin down the minimal, threshold conditions of
autonomous agency, we seem to come up against the conditions necessary for agency itself.
Consider, for example, an alleged paradigm case of an agent who fails to govern herself: a person who takes drugs
even though she would rather resist the motivating force of her addiction. It is widely agreed that, even if many people
who fit this description are merely weak-willed, not all of them are: some unwilling addicts are not self-governing in even
the minimal sense. According to coherentist and responsiveness-to-reasoning accounts, this is because such addicts are,
in effect, “passive bystanders” to their own motives. But even if we could find a satisfactory account of the relevant
passivity, this diagnosis would be problematic. For it assimilates the addict to someone whose behavior does not even
qualify as an action—someone, e.g., with Tourette’s Syndrome, whose verbal outbursts and bodily movements are not even voluntary. It thus fails to shed light on the conditions under which someone acts—intentionally, even deliberately—without being accountable for what she does. (The same problem arises for the proposal that autonomous agents differ from other agents in virtue of their special way of understanding what they are up to. Regardless of whether someone governs her intentional actions, she knows what she is doing without having to observe her behavior. If someone had to observe her own behavior in order to discover what she was up to, then this behavior would not qualify as her action.)

If someone’s motives directly defy her attempt to exercise authority over her actions, then their power is not only independent of her authority; they bypass her agency altogether. Even if under these conditions a person can acknowledge that there is something to be said for behaving as she does, she is as alienated from the power of her own motives as she is from the power of the physiological states that produce her reflex movements. She is not an autonomous agent because she is not an agent at all[18].

Of course, there is a sense in which the addict, but not the victim of Tourette’s, is responsive to reasons. But even if we could spell out this distinction in a satisfactory way, we still face the problem of explaining why an agent’s capacity to respond to external reality is relevant to her capacity to govern herself. And there are other problems too. If, for example, nonaction differs from action in virtue of being unresponsive to reasons, then it follows that nonrational animals never really act, or that acting from instinct is acting for reasons. It is not clear what considerations could overcome the implausibility of these implications.

4. Agents as Causes and the Practical Point of View

Even nonautonomous rational agents authorize the motivating power of the desires that move them to act. Yet there is an important sense in which the power behind their authorizations is not their own. What does the relevant impotence amount to? The incompatibilist, we saw, has a ready answer: an agent is prevented from exercising any power over her behavior when her authorization of this behavior can be traced to the determining influence of external powers. According to the incompatibilist, if the unwilling addict fails to govern herself, this is because her motives are determined by past states of affairs over which she does not have (and never did have) any control. And if a willing addict’s motives have a similar genesis, then she, too, is not accountable for their motivating power.

The familiar problem with this answer is that there seems to be no way for an agent to gain an extra measure of control over her motives simply by acquiring attitudes or judgments or other mental states that are not determined by anything else. If someone’s attitude toward her motives is not determined by any earlier state of affairs, then how can it be determined by her? This question pushes some with incompatibilist intuitions to attribute a special causal power to agents—a power of agency, which is not reducible to the effect one event has on another. On this view, a person can “agent-cause” a certain response to earlier events in a way that is not itself the effect of these earlier events. (Chisholm, Clarke, O’Connor, Taylor, R.) She can bring it about that she is motivated in a certain way, without anything determining her to do so. Thus, her power over her actions cannot be reduced to the power of external motivating forces.[19]

Such accounts of autonomous agency take seriously the need to distinguish an agent’s power from the power of the psychic forces that push and pull her. Yet the obscurities of the special sort of causation they evoke are enough to prevent most philosophers from embracing this conception of autonomous agency. To mention just a few familiar challenges: If agent-causing an event does not involve doing something to bring it about that the event occurs, then how does the agent exert her causal power? and why does this power produce its effects at one time rather than another? If, on the other hand, an agent must do something in order to agent-cause an action, then doesn’t this require that she undergo some change? and isn’t this change of state itself an event?

On the basis of these and other difficulties, many conclude that the appeal to agent-causation provides no more insight into autonomy than the simple assertion that we can sometimes govern the causal efficacy of our own motives. Incompatibilists who are sympathetic with this complaint nonetheless insist that even if an agent’s causal power must be understood in terms of the causal relations among events, autonomous agency is possible only if the conditions that immediately give rise to an agent’s intention need not suffice to produce it. If, they argue, an agent cannot authorize a given action without being determined to do so by powers beyond her control, then autonomous agency is an illusion. If our every action is an event in a deterministic causal chain, then we authorize our actions only in the sense that a figurehead authorizes the decisions she is forced to sign. Indeed, unlike a political puppet, we do not even have the option of defiance[20].

Others see things differently, however. They argue that the incompatibilist’s conclusion reflects a misunderstanding of the very nature of rational agency. In making their case, they take their lead from the philosopher who has contributed more than any other to our understanding of autonomy. Kant, they note, stresses the deep differences between the two points of view from which we can think about ourselves and our world.[21] We take up the theoretical point of view in order to gain knowledge about the nature of reality, and on this basis make predictions about which effects will follow from which causes. When we want to make up our minds about what to do, however, we take up the practical point of view. From this point of view, too, we survey the facts that are relevant to our decisions. But since none of these facts, taken singly or together, is intrinsically action-guiding, they cannot free us from the task of drawing our own conclusions about what we have reason to do. (Korsgaard, Bok)

This is true, the neo-Kantians point out, even if our decisions are the effect of causal powers over which we have no
control. Since no facts of which we could possibly become aware can force us to attribute any particular significance to them, and since in order to decide what to do, we must attribute some significance to the facts of which we are aware (even if this involves taking them to have no significance at all), our decisions are “up to us.” According to those who press this line of argument, our authority over our own actions would not be illusory even if our mode of exercising it were causally determined by events or states of affairs over which we have no control. Even under these circumstances, no events or states of affairs would have the power to determine what we have reason to do. So even under these circumstances, we have to reach our own conclusions on this point and acting accordingly. (Note that since this argument relies on the connection between determining what we have reason to do and determining what we will do, it can also be endorsed by those who take acting for reasons to involve responding to theoretical reasons. Indeed, there is a loose sense in which an agent’s point of view is “practical” as long as the truth of her predictions depends on the fact that she makes them.)

5. Conclusion

These last observations take us back to where we started. The political demand to be permitted to govern ourselves is grounded in the metaphysics of agency: any agent who faces the task of “making up her mind” has the authority to determine how she will act. On most occasions, what an agent does is the direct effect of her exercise of this authority. Yet there is also ample evidence that the capacity for self-government is vulnerable to any number of assaults; an agent’s authority over her actions is no guarantee that she has the power to determine how she exercises this authority. Agents can be deprived of their autonomy by brainwashing, depression, anxiety, fatigue; they can succumb to compulsions and addictions. To what, exactly, are we calling attention when we say that, under these conditions, an agent does not govern herself, even if she acts as she does because she thinks she has sufficient reason to do so, even if she has (thoroughly) considered the pros and cons of her options, and has endorsed her behavior on this basis, and even if she would have acted differently if there had been stronger reason to do so? Most agents who are capable of asking this question are confident that they are the authors of most of their actions, and are thus accountable for most of what they do. Nonetheless, as this brief survey indicates, the self-relation they thereby attribute to themselves is extremely difficult to pin down.[22]

Bibliography


Academic Tools

- How to cite this entry.
- Preview the PDF version of this entry at the Friends of the SEP Society.
- Look up this entry topic at the Indiana Philosophy Ontology Project (InPhO).
- Enhanced bibliography for this entry at PhilPapers, with links to its database.

Other Internet Resources

[Please contact the author with suggestions.]

Related Entries

- action
- agency
- Aristotle
- authority
- autonomy: in moral and political philosophy
- causation: the metaphysics of
- compatibilism
- determinism: causal
- feminist philosophy, topics: perspectives on autonomy
- feminist philosophy, topics: perspectives on the self
- free will
- Hume, David: on free will
- incompatibilism: (nondeterministic) theories of free will
- incompatibilism: arguments for
- intention
- Kant, Immanuel
- Kant, Immanuel: moral philosophy
- mental causation
- moral motivation
- moral responsibility
- Plato: ethics and politics in *The Republic*
- practical reason
- practical reason: and the structure of actions
- Reid, Thomas
- Sartre, Jean-Paul
- Schopenhauer, Arthur
- sovereignty
- weakness of will

Acknowledgments

I am grateful to Sven Nyholm for prodding me to see ways in which I could improve the entry when I was revising it for the first time. During that revision, he also worked tirelessly to help me track down references and input all of the new material.

Copyright © 2018 by
Sarah Buss <sbuss@umich.edu>
Andrea Westlund <westlund@uwm.edu>

Open access to the SEP is made possible by a world-wide funding initiative.
The Encyclopedia Now Needs Your Support
Please Read How You Can Help Keep the Encyclopedia Free
Personal autonomy is the capacity to decide for oneself and pursue a course of action in one's life, often regardless of any particular moral content. Political autonomy is the property of having one's decisions respected, honored, and heeded within a political context. Another distinction that can be made is between autonomy as a bare capacity to make decisions and of autonomy as an ideal. When autonomy functions as an ideal English examples for "personal autonomy" - More specifically, each person needs to have both physical health and personal autonomy. Since they have more wealth and privacy, the upper class has more personal autonomy than the rest of the population. This attempt requires the ability to participate in the societal setting in which an individual lives. More specifically, each person needs to have both physical health and personal autonomy.